DOCUMENT RESUME

ED 184 724 PS 011 374

AUTHOR MacPhee, David
TITLE Evidence for Reliable and Valid Dimensions of Infant Temperament.
PUB DATE Apr 80
NOTE 13p.; Paper presented at the Biennial Southeastern Conference on Human Development (6th, Alexandria, VA, April 17-19, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Age Differences; *Comparative Analysis; Factor Structure; Individual Differences; *Infants; Personality; *Personality Traits; Sex Differences; Social Differences; *Test Reliability; *Tests; *Test Validity
IDENTIFIERS Bayley Scales of Infant Development; EASI (Buss and Plomin); Infant Temperament Questionnaire; Temperament

ABSTRACT

The reliability and validity of three measures of infant temperament were compared in this study. The measures included the revised Carey (1978) Infant Temperament Questionnaire, a version of the Bayley (1969) Infant Behavior Record revised for completion by the parent, and a modified version of Buss and Plomin's (1975) EASI, an acronym standing for its four subscales--Emotionality, Activity level, Sociability and Impulsivity. Sixty-six mothers of healthy infants were asked to complete the battery of three infant temperament scales and a random subsample of mothers was asked to take a retest 21 days after the initial testing. Results showed that subscales from all three measures were correlated in meaningful ways, suggesting that there are independent, salient characteristics of the infant that can be measured by quite different rating scales. The strengths and weaknesses of each measure were identified and discussed. (JMB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED184724

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

EVIDENCE FOR RELIABLE AND VALID DIMENSIONS OF INFANT TEMPERAMENT

David MacPhee

Frank Porter Graham
Child Development Center
University of North Carolina
Chapel Hill, North Carolina 27514

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
David MacPhee

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the Sixth Biennial Southeastern Conference on Human Development, Alexandria, Va., April, 1980.

PS011374

The measurement of infant temperament has commanded more attention these days, now that the importance of individual differences in behavior and development is widely acknowledged. In this presentation, I will discuss two related issues pertaining to the assessment of temperament: One, "Which assessment techniques have most to recommend in terms of psychometric criteria?" and, two, "Which behaviors cluster into traits or dimensions of temperament?"

Several years ago, there would have been little need to address these issues because there was only one game in town, that of Thomas, Chess and Birch (1963; 1968; 1977). Recently, though, scales to measure infant temperament have been developed by Bates (1979), by Rothbart (Note 1), and by Buss and Plomin (1975). In addition, clinical tools such as Brazelton's Neonatal Exam (1973) and Bayley's Infant Behavior Record (1969) have gained popularity as measures of individual differences in infancy. The proliferation of these approaches to assessment raises some important questions, two of which I asked in this study. First, is one measure of infant temperament more reliable and valid than the others? These psychometric properties are an index of the usefulness of an instrument and, in part, determine the generalizability of findings. I hope to show that by comparing the reliability and validity of different temperament questionnaires, we may be able to select the best instrument for use in future research. Second, can we find a clearer definition of temperament, using multiple tests given to the same individuals? Subscales on different temperament questionnaires might be related to each other. Using convergent and divergent validity, evidence for a few meaningful dimensions of infant temperament might be found.

METHOD

I studied 66 mothers of healthy infants who were being seen for routine developmental examinations. All mother-infant pairs were recruited from well-child clinics in the Chapel Hill, North Carolina area. The sample was heterogeneous with respect to income and education -- social class ranged from poverty level to upper middle class -- and with respect to age of the mother, which ranged from less than 15 to 38 years of age ($\bar{X}$ = 25 years). The infants, half of whom were males, ranged in age from 3 to 9 months ($\bar{X}$ = 6 months).

I asked each mother to complete, at home, a battery of three infant temperament scales. These were given to the mothers at the clinic and were mailed back upon completion, with a 90% return rate. The measures included the revised Carey (1978) Infant Temperament Questionnaire, consisting of 95 items on 9 subscales; a version of the Bayley (1969) Infant Behavior Record revised for completion by the parent (instead of an examiner); and a modified version of Buss and Plomin's (1975) 30 item EASI, an acronym standing for its four subscales -- Emotionality, Activity level, Sociability and Impulsivity. Both the item format and content differed across measures, varying from concrete descriptions of specific behaviors and events (i.e. the Carey) to statements that were more projective and global in nature (the EASI and the Infant Behavior Record). In addition, the measures differed in the extent to which they categorize behaviors into dimensions of infant temperament.

RESULTS

I will present the results in two sections. The first will consider the reliability of each questionnaire while the second section will discuss the validity of each measure.

First, I analyzed the data for internal consistency, the results of which are shown in Table 1. Cronbach's alpha reliability (see Novick & Lewis, 1967), as it is called, is a variant of split half reliability which measures the cohesiveness of multi-item scales. As you can see, the EASI subscales, with a median coefficient of .72, are made up of more homogeneous items than the Carey subscales (median = .34). This finding may imply that some subscales of the Carey are not orthogonal dimensions of temperament but instead may overlap in a meaningful way. To test this, I factor analyzed the Carey subscale scores and found that the five subscales used by Thomas and his colleagues to classify infants as easy or difficult loaded on the same factor.

Next, I chose at random a subsample of mothers to receive a retest 21 days after initial testing; 80% of them returned the forms. As Table 1 shows, the coefficients were of modest size, though acceptable given the interval between tests. In addition, the Carey and the EASI do not differ in the magnitude of the test-retest correlations.

Turning to the evidence for concurrent validity, I completed a second order factor analysis in an effort to find dimensions of behavior common to all questionnaires. The variables included the 9 subscales of the Carey, the 4 subscales of Buss and Plomin's EASI, and factor scores from a four-factor solution to the Infant Behavior Record. The results, shown in Table 2, suggest that the three questionnaires share much in common. The Emotionality factor includes all five subscales of the Carey that are used to classify infants as easy or difficult plus the Emotionality subscale of the EASI. All three variables related to activity level appear on Factor II. Factors III and IV are not as clearly interpretable but seem to be related, first, to the ability of the infant to control arousal or impulses and, second, to the infant's response to tasks and objects.

The final set of analyses concerned individual and demographic differences in infant temperament. In these analyses, I used an easy-difficult summary variable in place of the five component subscales of the Carey. The use of this summary variable, which can be justified on the basis of past factor and item analyses, increased the power of the multivariate tests that were being used.

Using a multivariate analysis of variance, I looked at, in turn, sex, age, race and social class differences on the Carey and on the EASI. No sex differences were detected on either scale. Next, I examined age differences, the results of which are shown in Figure 1. Using SES as a covariate, I did find age differences on the Carey (multivariate test significant at the .05 level) but not on the EASI. As you can see, a linear trend was found on activity level (with older infants being rated as more active, $p < .001$) and on the easy-difficult dimension (with older infants being rated as more difficult, $p < .01$).

Finally, I looked at race and SES differences in ratings of infant temperament. Again, this was done by using a MANOVA with two factors (race and social class), covarying age. The results, illustrated in Figure 2, are similar to the age findings in that there were no differences on the EASI but there was a main effect on the Carey for social class (multivariate test significant at the .05 level). Significant univariate tests included distractability ($p < .05$), activity level ($p < .01$) and the easy-difficult variable ($p < .02$).

These results support two general conclusions, the first having to do with temperament traits and the second with the strengths and weaknesses of each questionnaire.

At the outset of this presentation, I noted that the measures I selected were quite diverse in content and style. Items on the Carey are concrete descriptions of everyday behaviors, with their frequency of occurrence rated on a 6 point scale. The Infant Behavior Record and the EASI, on the other hand,

ask the parent to make more global, impressionistic ratings of the infant's behavior. It is notable that subscales from all three measures are correlated in meaningful ways and, further, that these dimensions of infant behavior have appeared in various guises throughout the developmental literature. Thus, I would suggest that there are independent, salient characteristics of the infant that can be measured by quite different rating scales. Whether or not these dimensions are a product of parental perceptions or actual infant behavior awaits further observational study.

The second conclusion I would draw is that these measures of infant temperament have different strengths and weaknesses. The EASI, though less sensitive to demographic and individual differences, seems to have subscales that are more homogeneous. Furthermore, the Emotionality and Activity level subscales of the EASI had the highest loadings on their respective factors in the second order factor analysis. With 60% fewer items than the Carey, the EASI has much to recommend it.

We are still left with the significant age and social class differences on the Carey. Again, this raises the issue of the reliability of parental ratings. Do these differences in temperament scores arise in the parent or in the infant? Carey (Note 2) suggests that "researchers, in particular psychologists, tend to disbelieve reports by parents and deprecate them as mere 'parental perceptions'". It may be that the age differences in this study are real, since they replicate previous findings in the developmental literature. However, I would argue that the social class differences in ratings of infant temperament reflect different responses, by the parent, to the demands of the task. Middle class mothers are more familiar with test situations, have fewer difficulties in negotiating the language of rating scales, and may have a broader knowledge base to which they can refer in comparing

their infants to others'. Carey's standardization sample was largely white and middle class, and his norms are similar to ratings made by the same type of parent in this study. Therefore, it may be that rating scales are reliable in the hands of the middle class parent while additional methods may be warranted to supplement the ratings of lower SES parents.

In closing, I suggest that the study of infant temperament is still in its infancy. These findings imply that we need to know more about the variables that influence ratings of infant temperament. Of particular concern are differences in raters that may influence temperament scores, an issue that demands more observational work to verify parent ratings. Also, further psychometric studies, including standardization on different populations, are called for. Finally, we need to know more about age-related changes in temperament and how such changes fit into the broader picture of development.

REFERENCE NOTES

1. Rothbart, M. K., Furby, L., Kelley, S. R., and Hamilton, J. S. Development of a caretaker report temperament scale for use with 3-, 6-, 9-, and 12-month-old infants. Paper presented at the Biennial Meeting of the Society for Research in Child Development, New Orleans, March 1977.

2. Carey, W. B. The importance of temperament-environment interaction for child health and development. Paper presented at the Institute for the Study of Exceptional Children, Princeton, N. J., March 1979.

REFERENCES

Bates, J. E., Freeland, C. A. B. and Lounsbury, M. L. Measurement of infant difficultness. Child Development, 1979, 50, 794-803.

Bayley, N. Manual, Bayley Scales of Infant Development. New York: Psychological Corporation, 1969.

Brazelton, T. B. Neonatal Behavioral Assessment Scale. Clinics in Developmental Medicine, no. 50. Philadelphia: Lippincott, 1973.

Buss, A. H. and Plomin, R. A temperament theory of personality development. New York: Wiley, 1975.

Carey, W. B. and McDevitt, S. C. Revision of the Infant Temperament Questionnaire. Pediatrics, 1978, 61, 735-739.

Novick, M. R. and Lewis, C. Coefficient alpha and the reliability of composite measurements. Psychometrika, 1967, 32, 1-13.

Thomas, A., Chess, S., Birch, H. G., Herzig, M. and Korn, S. Behavioral individuality in early childhood. New York: New York University Press, 1963.

Thomas, A., Chess, S. and Birch, H. G. Temperament and behavior disorders in children. New York: New York University Press, 1968.

Thomas, A. and Chess, S. Temperament and development. New York: Bunner/Mazel, 1977.

TABLE 1

Internal consistency and test-retest reliability
for the EASI and Carey ITQ subscales

	Alpha coefficient (N=66)	Test-retest (N=27)
EASI		
Emotionality	.73	.59
Activity level	.76	.67
Sociability	.71	.68
Impulsivity	.69	.72
Carey ITQ		
Adaptability	.34	.76
Approach	.34	.76
Rhythmicity	.25*	.81
Intensity	.32**	.45*
Mood	NS	.79
Activity level	.37	.67
Distractability	NS	.70
Persistence	.41	.67
Threshold	.37	.65

NOTE: Unless indicated otherwise, all values are significant beyond the .005 level.

* $p < .05$
** $p < .01$

TABLE 2

Second order factor analysis of three temperament questionnaires[a]

Variables[b]		Loadings I	II	III	IV
I. Emotionality (46.7%):					
Emotionality	(EASI)	*.80*			
Approach	(ITQ)	*.80*			
Intensity	(ITQ)	*.78*			
Adaptability	(ITQ)	*.78*			
Mood	(ITQ)	*.75*			
Rhythmicity	(ITQ)	*.67*			
Cooperative	(IBR)	.49	.25	.30	-.25
II. Activity level (23.1%):					
Activity level	(EASI)		*.73*		
Activity level	(ITQ)		*.73*		
Activity level	(IBR)		*.65*		
Sociability	(EASI)		.47		.38
III. Reactivity (16.4%):					
Reactivity	(IBR)			*.76*	
Distractability	(ITQ)			*.64*	
Impulsivity	(EASI)			*.61*	
IV. Persistence (13.6%):					
Persistence	(ITQ)				*.73*
Threshold	(ITQ)				*.71*
Object orient.	(IBR)		.38		.53

[a] Percentage of common variance accounted for is in parentheses. Italicized loadings indicate those items which discriminate one factor from the others.

[b] ITQ - Carey Infant Temperament Questionnaire
EASI - Buss and Plomin EASI
IBR - Bayley Infant Behavior Record

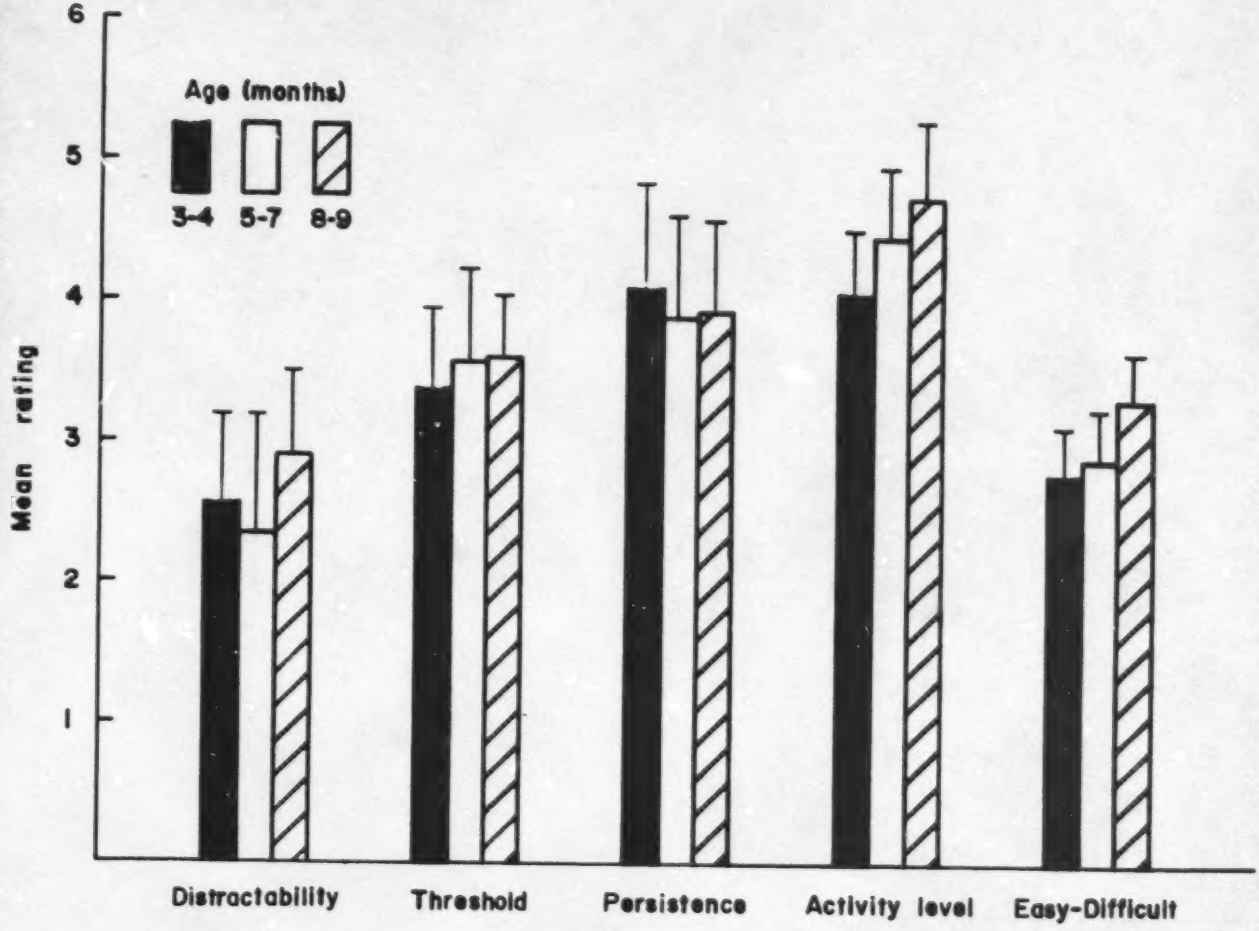

Figure 1: Means and standard deviations on the Carey Infant Temperament Questionnaire, by age.

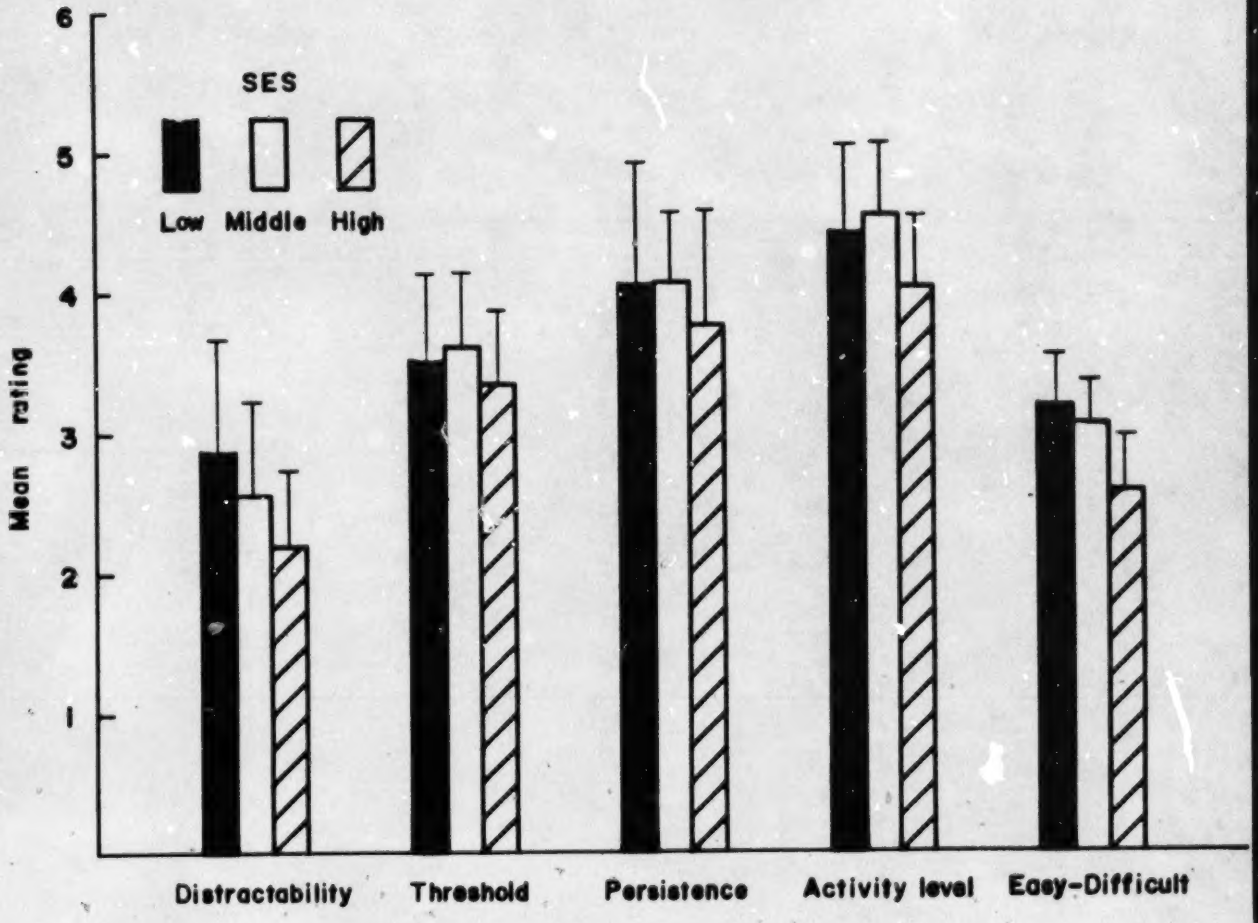

Figure 2: Means and standard deviations on the Carey Infant Temperament Questionnaire, by social class.